

**Kaufman -
Trading
Systems &
Methods -**

Appendix 2

VISIT...

LANZAROTE
Caliente.COM

APPENDIX 2

Method of Least Squares

The following programs, written in Microsoft FORTRAN for the IBM PC,' solve the twovariable regression analysis using the method of least squares. The program allows for the regression of two series against one another, and the regression of one series against time. OPERATING INSTRUCTIONS in the following example, the user's entries are underlined:

LSTSQ

Enter data in sets of 2 (Y then X).

If second variable is omitted, sequential numbers will be used for x.

Extra <return> ends input.

Enter below:

YYYYYY XXXXXX

1.27 2.43

1.19 2.26

1.10 2.15

<RETURN>

Print data (Y/N)? Y

(Computer prints data and solutions)

Sample results (Y/N)?Y

Enter start, end and increments below

sssss eeeee iiii

1.00 7.00 .50

(Computer prints sample results)

Plot results (Y/N)?Y

Name for X-scale (9 chars)?

SOYBEANS

Name for Y-scale (9 chars)?

CORN

Which plot?

- 1: Linear
- 2: Logarithmic
- 3: Exponential
- 4: Curvilinear

Which number? 1

(Computer plots)

Another plot (Y/N)?Y

Which plot?

- 1: Linear

(and so forth)

COMPUTER PROGRAMS

1 \$TITLE: 'LSTSQ: LEAST SQUARES'

2 \$SUBTITLE: 'COPYRIGHT 1986 P J KAUFMAN'

' BASIC programs for both plotting and least-squares regression can be found in F.R. Ruckdeschel, *BASIC Scientific Subroutines, Volume I* (Byte/McGraw-Hill, Peterborough, NH, 1981).

```

3 $STORAGE:2
4   PROGRAM LSTSQ
5 C---- Least Squares Regression Analysis for 2 variables
6
7   CHARACTER*1 ANS
8   DIMENSION X(100),Y(100)
9
10  DATA MAX/100/
11
12  OPEN(6,FILE='PRN')
13
14 C---- Read data in sets of 2 variables
15   I = 1
16   WRITE(*,7000)
17 7000 FORMAT(' Enter data in sets of 2 (Y then X).'/
18   +         ' If second variable is omitted, sequential numbers',
19   +         ' will be used for X.'/' Extra <return> ends input.'/'
20   +         ' Enter below:/'
21   +         ' YYYYYY XXXXXX')
22 10 READ(*,5000)Y(I),X(I)
23 5000 FORMAT(BN,F6.2,F7.2)
24 IF(Y(I).EQ.0)GOTO 30
25 IF(X(I).EQ.0)X(I)=I
26 IF(I.EQ.MAX)THEN
27   WRITE(*,7001)MAX
28 7001 FORMAT(' Data input reached max of',I5,'. Processing begins.')

```

1 $TITLE: 'RA2V: Regression for 2 Variables'
2 $SUBTITLE: 'Copyright 1986 PJ Kaufman'
3 $STORAGE:2
4 SUBROUTINE RA2V(Y,X,N)
5
6 C---- Regression Analysis for 2 Variables
7 C---- With Linear, Power and Log Transformations
8
9 C---- Y = Dependent variable
10 C---- X = Independent variable
11 C---- N = Number of data entries

```


```

```

12
13 CHARACTER*1 ANS
14 CHARACTER*2 SIGN,SIGNC
15 CHARACTER*9 YNAME,XNAME
16
17 DIMENSION X(100),Y(100),ZA(4),ZB(4),ZR(4)
18
19 DATA MAX/100/
20
21 C---- Perform linear regression
22 ITYPE=1
23 CALL BSTFIT(Y,X,N,ZA(1),ZB(1),ZR(1),SD,ITYPE)
24 SIGN=' '
25 IF(ZB(1).GE.0)SIGN=' +'
26
27 WRITE(6,6000)ZA(1),SIGN,ZB(1),ZR(1),SD
28 6000 FORMAT(/54X, 'r2 St dev'//
29      + ' LINEAR Y=',F8.3,A2,F8.3,'*X',18X,F7.1,F9.3)
30
31 CALL LSTSQ2(Y,X,N,ZA(4),ZB(4),C,S4)
32
33 SIGN=' '
34 IF(ZB(4).GE.0)SIGN=' +'
35 SIGNC=' '
36 IF(C.GE.0)SIGNC=' +'
37 ITYPE = 4
38
39 WRITE(6,6004)ZA(4),SIGN,ZB(4),SIGNC,C,S4
40 6004 FORMAT(/' CURVI Y=',F8.3,A2,F8.3,'*X',A2,F8.3,'*X2',12X,F9.3)
41
42 ITYPE=2
43 CALL BSTFIT(Y,X,N,ZA(2),ZB(2),ZR(2),SD,ITYPE)
44 SIGN=' '
45 IF(ZB(2).GE.0)SIGN=' +'
46
47 WRITE(6,6001)ZA(2),SIGN,ZB(2),ZR(2),SD
48 6001 FORMAT(/' LOG Ln(Y)=',F8.3,A2,F8.3,'*Ln(X)',10X,F7.1,F9.3)
49
50 ITYPE=3
51 CALL BSTFIT(Y,X,N,ZA(3),ZB(3),ZR(3),SD,ITYPE)
52 SIGN=' '
53 IF(ZB(3).GE.0)SIGN=' +'
54
55 WRITE(6,6002)ZA(3),SIGN,ZB(3),ZR(3),SD
56 6002 FORMAT(/' EXPON Ln(Y)=',F10.3,A2,F10.3,'*X',10X,F7.1,F9.3)
57
58 WRITE(*,7000)
59 7000 FORMAT('Sample results (Y/N)?'\)
60 READ(*,5000)ANS
61 5000 FORMAT(A1)
62 IF(ANS.NE.'N')THEN
63     WRITE(*,7001)
64 7001 FORMAT(/' Enter start, end and increments below'/
65      + 'sssss eeeee iiii')
66 READ(*,5001)P,END,STEPS
67 5001 FORMAT(BN,F6.2,2F7.2)
68
69 WRITE(6,6009)
70 6009 FORMAT(/12X, 'X LINEAR CURVI LOG EXPON'//)
71
72 C---- Calculate all four fits at the same time

```

```

73 100    Y1 = ZA(1) + ZB(1)*P
74        Y2 = ZA(4) + ZB(4)*P + C*P*P
75        Y3 = ZA(2) + ZB(2)*ALOG(P)
76        Y3 = EXP(Y3)
77        Y4 = ZA(3) + ZB(3)*P
78        Y4 = EXP(Y4)
79
80        WRITE(6,6010)P,Y1,Y2,Y3,Y4
81 6010    FORMAT(5X,5F8.2)
82
83        IF(P.LT.END)THEN
84            P = P + STEPS
85            GOTO 100
86        ENDIF
87
88    ENDIF
89
90    WRITE(*,7002)
91 7002    FORMAT(/' Plot results (Y/N)?'\')
92    READ(*,5000)ANS
93    IF(ANS.NE.'N')THEN
94        WRITE(*,7003)
95 7003    FORMAT(' Name for x-scale (9 chars)?')
96        READ(*,5002)XNAME
97 5002    FORMAT(A)
98        WRITE(*,7004)
99 7004    FORMAT(' Name for y-scale (9 chars)?')
100        READ(*,5003)YNAME
101 5003    FORMAT(A9)
102 120    WRITE(*,7005)
103 7005    FORMAT(/' Which plot?/' 1: Linear/' 2: Logarithmic/'
104 +        ' 3: Exponential/' 4: Curvilinear/'
105 +        ' Which number?'\')
106        READ(*,5004)I
107 5004    FORMAT(BN,I4)
108        TC = C
109        IF(I.NE.4)TC = 0
110        CALL PLOTTRA(N,X,Y,ZA(I),ZB(I),TC,I,XNAME,YNAME)
111        WRITE(*,7006)
112 7006    FORMAT(/' Another plot (Y/N)?'\')
113        READ(*,5000)ANS
114        IF(ANS.NE.'N')GOTO 120
115    ENDIF
116
117    RETURN
118    END

1 $TITLE: 'LSTSQ2: 2nd Order Least Squares'
2 $SUBTITLE: 'P.J. Kaufman'
3 $STORAGE:2
4     SUBROUTINE LSTSQ2(Y,X,N,A,B,C,S)
5
6 C---- Second Order (Curvilinear) Regression for 2 variables
7
8 C---- Y = Dependent variable
9 C---- X = Independent variable
10 C---- N = Number of data entries
11 C---- A = Constant term
12 C---- B = Coefficient of X
13 C---- C = Coefficient of X**2
14 C---- S = Standard deviation of residuals

```

```

15
16     DIMENSION X(2),Y(2)
17
18     DOUBLE PRECISION SXX,SXY,SYY,SXX2,SX2X2,SYX2
19
20 C---- Find the mean of x and y
21     AX = 0
22     AY = 0
23     DO 10 I = 1,N
24         AX = AX + X(I)
25     10  AY = AY + Y(I)
26     AX = AX/N
27     AY = AY/N
28
29 C---- Initialize for summations
30     SXX = 0
31     SXY = 0
32     SYY = 0
33     SXX2 = 0
34     SX2X2 = 0
35     SYX2 = 0
36
37 C---- Sums
38     DO 20 I = 1,N
39         XI = X(I)
40         YI = Y(I)
41         XM = XI - AX
42         YM = YI - AY
43         XM2 = XI*XI - AX*AX
44         SXX = SXX + XM*XM
45         SXY = SXY + XM*YM
46         SYY = SYY + YM*YM
47         SXX2 = SXX2 + XM*XM2
48         SX2X2 = SX2X2 + XM2*XM2
49     20  SYX2 = SYX2 + YM*XM2
50
51     B = (SXY*SX2X2 - SYX2*SXX2) / (SXX*SX2X2 - SXX2*SXX2)
52     C = (SXX*SYX2 - SX2X2*SXY) / (SXX*SX2X2 - SXX2*SXX2)
53     A = AY - B*AX - C*AX*AX
54
55 C---- Standard deviation of residuals
56     S = 0
57     DO 150 J = 1,N
58         TY = Y(J)
59         TX = X(J)
60         RV = A + B*TX + C*TX*TX
61         ERR = TY - RV
62         AVGY=AVGY+TY
63     150  S = S + ERR*ERR
64
65     S = SQRT(S/(N-1))
66
67     RETURN
68     END

```

```

1 $TITLE: 'REGVAL: Get regression value'
2 $SUBTITLE: 'Copyright 1986 PJ Kaufman'
3 $STORAGE:2
4     FUNCTION REGVAL(X,A,B,ITYPE)
5

```

```

6 C---- Transform data from log form to normal form based upon the
7 C---- type of regression fit (transmitted via ITYPE)
8
9      GOTO (110,120,130),ITYPE
10
11 C---- Linear
12 110  TY=A+B*X
13      GOTO 140
14
15 C---- Log
16 120  TY=A+B*ALOG(X)
17      TY=EXP(TY)
18      GOTO 140
19
20 C---- Exponential
21 130  TY=A+B*X
22      TY = EXP(TY)
23
24 140  REGVAL=TY
25
26      RETURN
27      END

```

119

```

120      SUBROUTINE BSTFIT(Y,X,N,A,B,R,SD,ITYPE)
121
122 C---- Basic linear regression
123 C---- N      Number of input items
124 C---- X      Independent variable
125 C---- Y      Dependent variable
126 C---- A      Y-intercept
127 C---- B      Slope
128 C---- R      Correlation coefficient
129 C---- SD     Standard deviation of residuals
130 C----
131 C---- If ITYPE = 1 then linear (no prior x,y conversion)
132 C----           = 2 then log (x=log x,y=log y)
133 C----           = 3 then exponential (x=log x)
134
135      DIMENSION X(2), Y(2)
136
137      DOUBLE PRECISION SUMX2,SUMY2,XY
138
139      SUMX=0.
140      SUMX2=0.
141      SUMY=0.
142      SUMY2=0.
143      XY=0.
144
145      DO 100 I=1,N
146          TX=X(I)
147          IF(ITYPE.EQ.2)TX=ALOG(TX)
148          TY=Y(I)
149          IF(ITYPE.NE.1)TY=ALOG(TY)
150          SUMX=SUMX+TX
151          SUMX2=SUMX2+TX*TX
152          SUMY=SUMY+TY
153          SUMY2=SUMY2+TY*TY
154          XY=XY+TX*TY
155 100  CONTINUE

```



```

156      IF(N.LT.2)RETURN
157      B=(N*XY-SUMX*SUMY)/(N*SUX2-SUMX*SUX)
158      A=(SUMY-B*SUX)/N
159
160
161 C---- Correlation coefficient
162      R1=N*XY-SUMX*SUMY
163      R2=(N*SUX2-SUMX*SUX)*(N*SUY2-SUMY*SUMY)
164      R = ((R1**2)/R2)*100
165
166 C---- Standard deviation around fitted function
167      SD=0
168      DO 150 J=1,N
169          TY=Y(J)
170          TX=X(J)
171          RV=REGVAL(TX,A,B,ITYPE)
172          T=TY-RV
173 150      SD=SD+T*T
174      SD=SQRT(SD/(N-1))
175
176      RETURN
177      END

```

LEAST-SQUARES SOLUTION FOR CORN VERSUS SOYBEANS

Least-Squares Regression Analysis Data Input

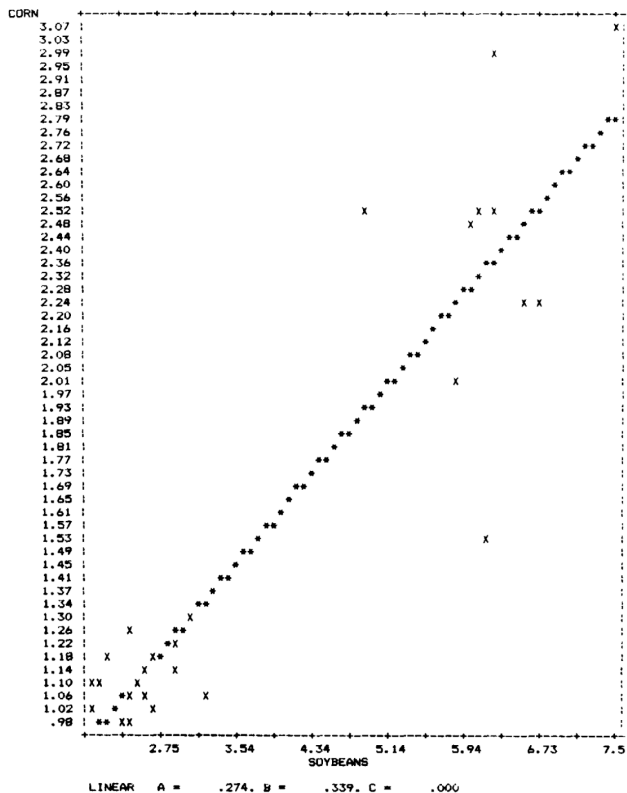
Y (dep) X (ind)

1.27	2.43
1.19	2.26
1.10	2.15
1.10	2.07
1.05	2.03
1.00	2.45
.98	2.36
1.09	2.44
1.12	2.52
1.18	2.74
1.16	2.98
1.24	2.93
1.03	2.69
1.08	2.63
1.15	2.63
1.33	3.08
1.08	3.24
1.57	6.22
2.55	6.12
3.02	6.33
2.54	4.92
2.25	6.81
2.02	5.88
2.25	6.61
2.52	6.28
3.11	7.61
2.50	6.05

			r ²	St dev
LINEAR	Y=	.274 + .339*X	84.2	.278
CURVI	Y=	.310 + .323*X + .002*X ²		.278
LOG	Ln(Y)=	-.616 + .799*Ln(X)	86.3	.282
EXPON	Ln(Y)=	-.373 + .195*X	87.3	.283

X	LINEAR	CURVI	LOG	EXPON
1.00	.61	.63	.54	.84
1.50	.78	.80	.75	.92
2.00	.95	.96	.94	1.02
2.50	1.12	1.13	1.12	1.12
3.00	1.29	1.29	1.30	1.24
3.50	1.46	1.46	1.47	1.36
4.00	1.63	1.63	1.64	1.50
4.50	1.80	1.80	1.80	1.66
5.00	1.97	1.97	1.95	1.83
5.50	2.14	2.14	2.11	2.01
6.00	2.31	2.31	2.26	2.22
6.50	2.48	2.48	2.41	2.45
7.00	2.65	2.66	2.56	2.70

Plot of Regression Analysis

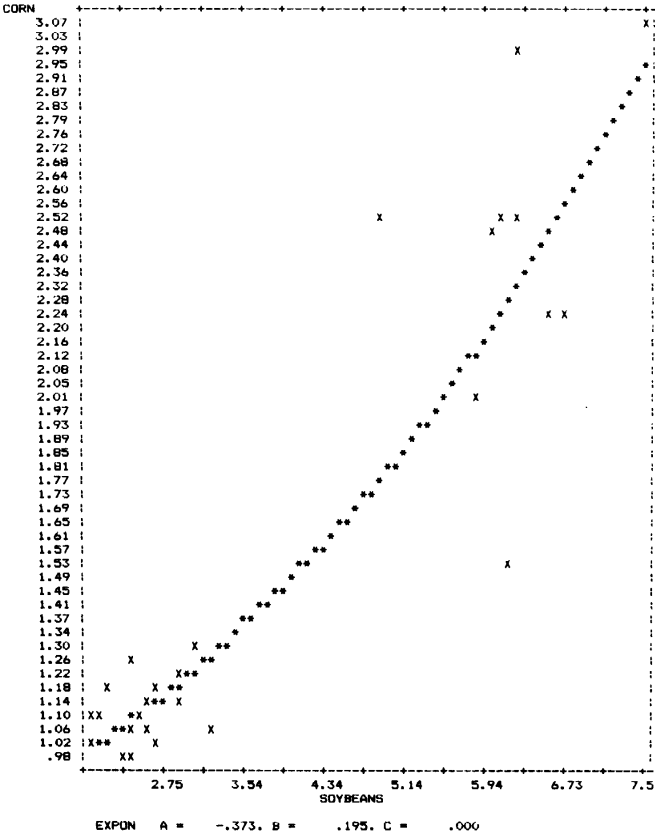


Scatter plot showing the relationship between Soybeans yield (kg/ha) on the x-axis and Corn yield (kg/ha) on the y-axis for the 1980-81 season. The x-axis ranges from 2.75 to 7.5 kg/ha, and the y-axis ranges from 1.02 to 3.07 kg/ha. Data points are marked with 'X' and '*' symbols. A dashed line represents the 1:1 relationship.

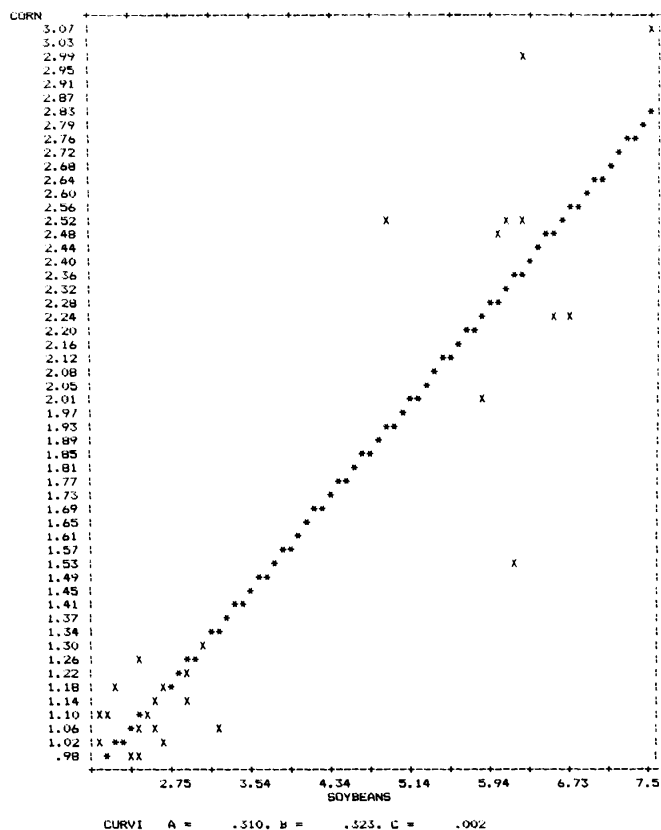
LOG A = -.616. B = .799. C = .000

643

Plot of Regression Analysis



Plot of Regression Analysis



LEAST-SQUARES SOLUTION FOR SOYBEANS ONLY

Least Squares Regression Analysis

Data Input

Y (dep)	X (ind)
2.43	1.00
2.26	2.00
2.15	3.00
2.07	4.00
2.03	5.00
2.45	6.00
2.36	7.00
2.44	8.00
2.52	9.00
2.74	10.00
2.98	11.00
2.93	12.00
2.69	13.00
2.63	14.00
2.63	15.00
3.08	16.00
3.24	17.00
6.22	18.00
6.12	19.00
6.33	20.00
4.92	21.00
6.81	22.00
5.88	23.00
6.61	24.00
6.28	25.00
7.61	26.00
6.05	27.00

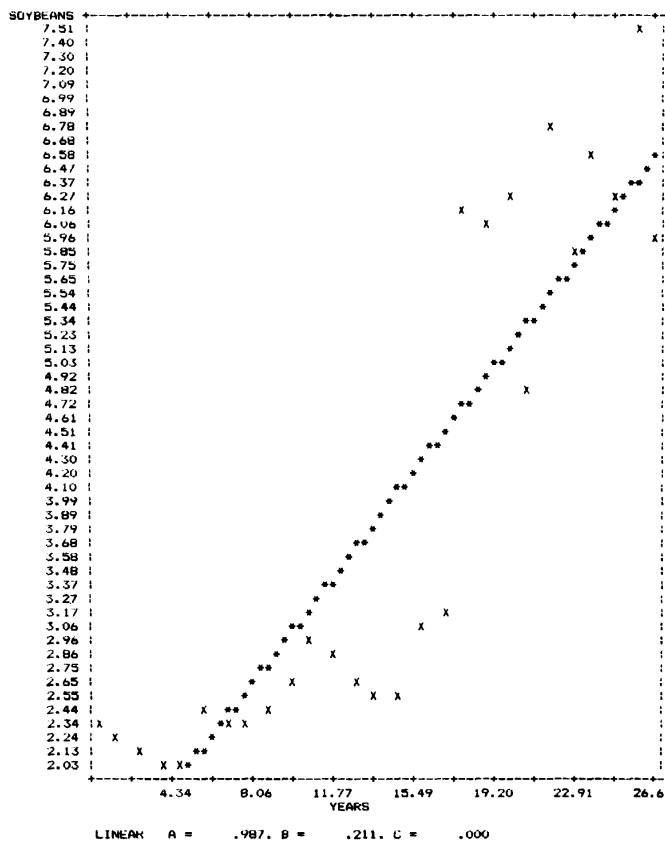
LINEAR	Y=	.987 +	.211*X	r2	St dev
CURVI	Y=	1.684 +	.112*X + .004*X2	78.3	.881
LOG	Ln(Y)=	.271 +	.416*Ln(X)	58.0	1.212
EXPON	Ln(Y)=	.527 +	.053*X	83.1	.806

X	LINEAR	CURVI	LOG	EXPON
1.00	1.20	1.80	1.31	1.79
2.00	1.41	1.92	1.75	1.88
3.00	1.62	2.05	2.07	1.99
4.00	1.83	2.19	2.34	2.09
5.00	2.04	2.33	2.56	2.21
6.00	2.25	2.48	2.77	2.33
7.00	2.47	2.64	2.95	2.45
8.00	2.68	2.80	3.12	2.59
9.00	2.89	2.98	3.27	2.73
10.00	3.10	3.16	3.42	2.87
11.00	3.31	3.34	3.56	3.03
12.00	3.52	3.53	3.69	3.19
13.00	3.73	3.74	3.82	3.37
14.00	3.94	3.94	3.93	3.55
15.00	4.15	4.16	4.05	3.74
16.00	4.37	4.38	4.16	3.94
17.00	4.58	4.61	4.27	4.16
18.00	4.79	4.84	4.37	4.38
19.00	5.00	5.09	4.47	4.62
20.00	5.21	5.34	4.56	4.87
21.00	5.42	5.59	4.66	5.14

22.00	5.63	5.86	4.75	5.41
23.00	5.84	6.13	4.84	5.71
24.00	6.05	6.41	4.92	6.02
25.00	6.27	6.70	5.01	6.34
26.00	6.48	6.99	5.09	6.69
27.00	6.69	7.29	5.17	7.05
28.00	6.90	7.60	5.25	7.43
29.00	7.11	7.91	5.33	7.83
30.00	7.32	8.23	5.40	8.26
31.00	7.53	8.56	5.48	8.71
32.00	7.74	8.89	5.55	9.18
33.00	7.95	9.24	5.62	9.68
34.00	8.17	9.59	5.69	10.20
35.00	8.38	9.94	5.76	10.75
36.00	8.59	10.31	5.83	11.34
37.00	8.80	10.68	5.90	11.95
38.00	9.01	11.06	5.96	12.60
39.00	9.22	11.44	6.03	13.28
40.00	9.43	11.83	6.09	14.00

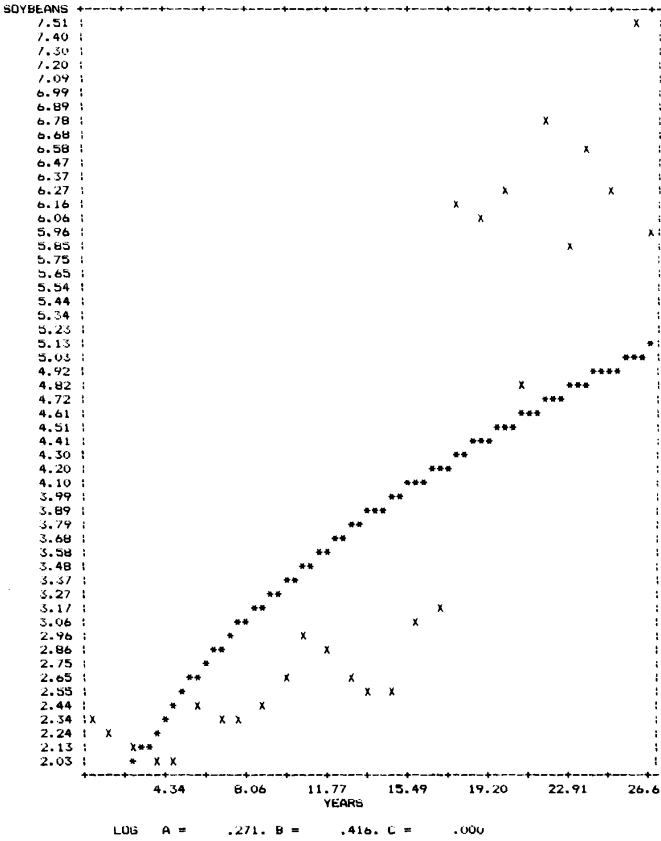
647

Plot of Regression Analysis

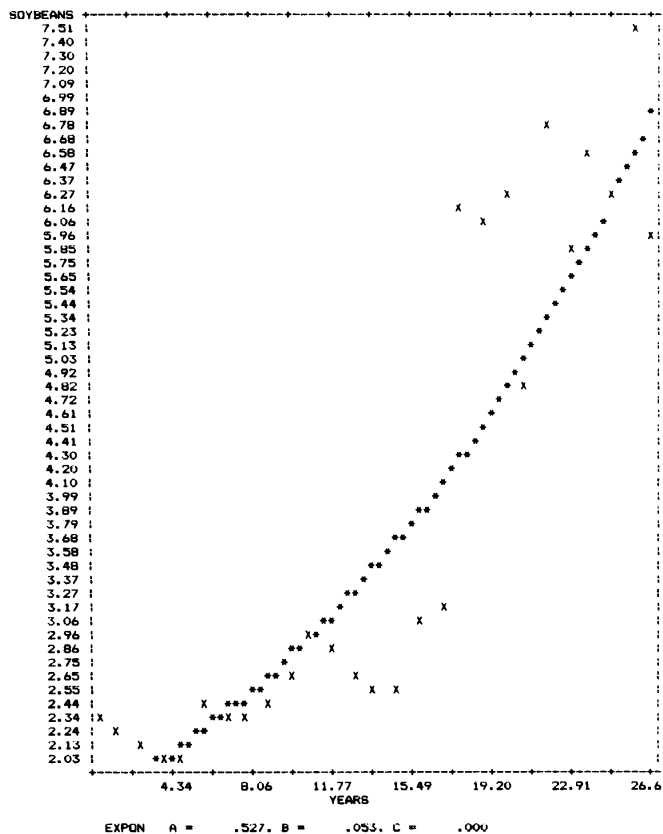


648

Plot of Regression Analysis



Plot of Regression Analysis



Plot of Regression Analysis

